TUTORIAL TUTORIAL TUTORIAL TUTORIAL TUTORIAL TUTORIAL TUTORIAL TUTORIAL

Pitfalls of Accelerated Testing

William Q. Meeker
Iowa State University, Ames

Luis A. Escobar
Louisiana State University, Baton Rouge

Key Words — Arrhenius, Censored data, Degradation, Masked failure mode.

Summary & Conclusions — Accelerated tests are used to obtain timely information on product-life or performance-degradation over time. Test units are used more frequently than usual or are subjected to higher than usual levels of accelerating variables like temperature & voltage. Then the results are used, through an appropriate physically-based statistical model, to make predictions about product life or performance over time, at the more moderate use-conditions. The extrapolative predictions inherent in the use of accelerated testing raise serious concerns, and the use of accelerated testing has many dangerous pitfalls. This paper warns potential users about some of these pitfalls.

1. INTRODUCTION

Acronyms[1]

ALT	accelerated life test(ing)
IC	integrated circuit
ML	maximum likelihood
UV	ultra-violet

ALT can be useful for obtaining timely information about materials & products. Refs [5 – 7], *etc* describe applications, models, statistical methods, and examples. There are, however, many important potential pitfalls that could cause an ALT to lead to seriously incorrect conclusions. Users of ALT should be careful to avoid these pitfalls.

2. PITFALL 1

Multiple (Unrecognized) Failure Modes

High levels of accelerated variables or stresses, like temperature or voltage, can induce failure modes that would not be observed at usual operating conditions In some cases, new failure modes result from a fundamental change in the mechanism causing the material or component to degrade or fail. For example, instead of simply accelerating a failure-causing chemical process, increased temperature can actually change certain material properties, *eg*, cause melting. In less extreme cases, high levels of an accelerating variable changes the relationship between life and the variable, *eg*, life at high temperatures might not be linear in inverse absolute temperature, as predicted by the widely-used Arrhenius relationship. This can result when different underlying failure-causing mechanisms are affected differently by temperature, *eg*, when different chemical reactions have importantly different activation energies.

If new, different failure modes arise at higher levels of an accelerated variable, and if the new failure modes are recognized, the new failure modes can, in some situations, be accounted for in the data analysis. If the potential failure times for the different failure modes are *s*-independent, *eg*, because they are caused by different failure mechanisms, one can treat the failure time for the new failure modes as censored observations. Ref [6: chapter 7] gives several examples. In such cases, however, the resulting censoring can severely limit the information available on the failure mode of interest. If other failure modes are present but not recognized in data analysis, seriously incorrect conclusions are possible, especially if the different failure modes are accelerated differently, *eg*, temperature affects one failure mechanism much differently than the others.

3. PITFALL 2

Failure To Quantify Uncertainty Properly

It is important to recognize that there is uncertainty in statistical estimates. Basing decisions on point estimates alone can, in many applications, be seriously misleading. Standard *s*-confidence bounds quantify uncertainty arising from limited data and reflecting knowledge of assumed inputs.

For example, [5: chapter 19] describes the analysis of accelerated life test data for a new-technology IC device. Figure 1 is a lognormal probability plot reflecting: a) failures observed at 250°C and 300°C (no failures had been observed at 150°C, 175°C, or 200°C), b) an estimate of $F(t)$, the lognormal Cdf, and c) a set of approximate 95% *s*-confidence intervals for $F(t)$. For this discussion we ignore the indication that the lognormal shape parameter might have changed from 250°C to 300°C; but see [5: sections 19.3.2 & 22.2] for further discussion of this

[1]The singular & plural of an acronym are always spelled the same.

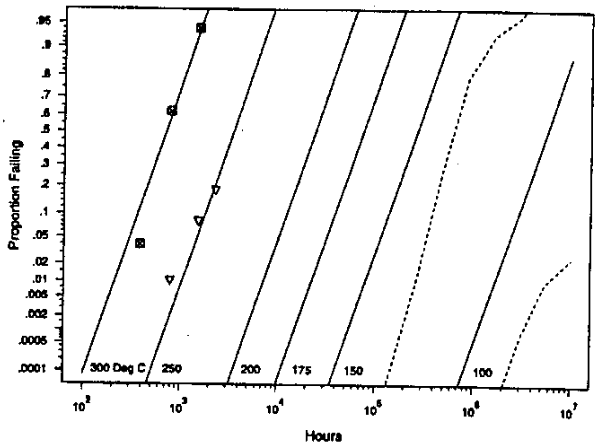

Figure 1: Lognormal Probability Plot #1 for a New-Technology IC

These curves/plots show the ML estimates for $F(t)$ along with a set of approximate 95% *s*-confidence intervals (dotted line) for $F(t)$ at 100°C based on the Arrhenius-lognormal model.

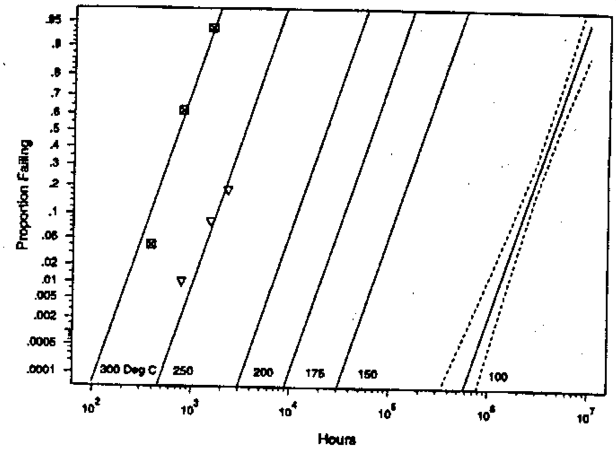

Figure 2: Lognormal Probability Plot #2 for a New-Technology IC

These curves/plots show the Arrhenius-lognormal model ML estimates and a set of approximate 95% *s*-confidence intervals for $F(t)$ at 100°C, with given $E_a = 0.8$ eV.

point. The *s*-confidence intervals in figure 1 indicate an enormous amount of uncertainty in life at 100°C, due to the: a) small sample size, b) small number of failures, and c) large amount of extrapolation in temperature when activation energy is estimated from the available data.

The corresponding analysis shown in figure 2 uses a given value of activation energy for the life-temperature relationship. Because the activation energy is not known exactly, the precision exhibited in this plot is too tight. For many applications, neither of these extremes could provide a proper quantification of uncertainty. A Bayes[2] analysis that uses a probability distribution to quantify the uncertainty in prior information would provide an appropriate compromise for situations where there is useful well-know prior information about activation energy [5: chapter 22]. Such information could be available from careful experience with the same failure mechanism.

It is important to remember that *s*-confidence bounds do not account for model uncertainty (which can be tremendously amplified by extrapolation in accelerated testing). In general, performing sensitivity analysis is very important in any quantitative analysis involving uncertainty and is particularly useful for assessing the effects of model uncertainty. For example, one can re-run analyses under different assumed models to see the effect that different model assumptions have on bottom-line conclusions.

[2]The reader is reminded of the philosophical differences between classical & Bayes probability theory. In the former, the probability is used to model the relative frequency of an event, whereas in the latter, the probability is used to model the user's degree-of-belief that the event will occur. Thus, in a classical interval estimate, the *s*-confidence level relates to the frequency with which the estimator produces intervals that cover the true mean, while in the Bayes paradigm, the *b*-credibility level relates to the user's degree-of-belief that the true mean lies in the interval.

4. PITFALL 3

Multiple Time-Scales and Multiple Factors Affecting Degradation

In any life-data reliability-analysis problem it is important to consider carefully the appropriate time-scale to use for the analysis. See Ref [5: section 1.3.4] for further discussion & references. These issues become even more important with accelerated testing. Common acceleration-methods might not accelerate all time scales simultaneously. For example, in an accelerated test to estimate the life-time characteristics of a composite material, chemical degradation over time changes the material ductility. Failures, however, are actually caused by stress cycles during use, leading to initiation and growth of cracks. But the effect of cycling depends on the material ductility.

An incandescent light bulb generally fails when its filament breaks. During burn time the bulb's filament goes through a sublimation process, eventually leading to failure. There are, however, other factors that can shorten a bulb's life. In particular, on-off cycles can induce both thermal & mechanical shocks that can lead to fatigue cracks in the filament. Thus the on-off frequency also affects bulb life. Accelerating only the burn time, *eg*, by testing at higher voltage, might give misleading predictions of life in an environment with many on-off cycles. Relatedly, light bulbs operated in an environment with physical vibration, *eg*, in automobiles, on large ships, or in a motorized appliance, often exhibit shorter lives — depending on the frequency & amplitude of the vibrations as well as the bulb's design.

The degradation of paints & coatings depend on several factors relating to time scales. Most coatings degrade chemically over time. UV light accelerates the degradation process of many kinds of coatings, as does temperature. The number of wet-dry cycles is also important to coating life, but generally relates to a separation or peeling failure-mode that is different from (but perhaps related to) the chemical degradation mechanism. Each of these experimental variables, and each failure mode, has its own underlying time-scale. Similarly, there is a mixture of use conditions in the field, *eg*, some automobiles are driven in the north and some in the south; some spend substantial time in direct sunlight, others do not.

In simple accelerated tests, where the distribution of use conditions is given, one might be able to accelerate time scales in appropriate proportion(s). In other applications, it is necessary to use a multiple-variable accelerated test, *eg*, varying temperature, humidity, and UV exposure. What is really needed (but usually difficult or expensive to determine in practice) is an adequate physical model to describe the relationship among these variables and degradation and life.

5. PITFALL 4

Masked Failure-Mode

Figure 3 shows a graph of what might illustrate the results of a typical accelerated life test if there were just a single failure mode and if increased temperature accelerated that failure mode in a simple manner, described by the Arrhenius relationship.

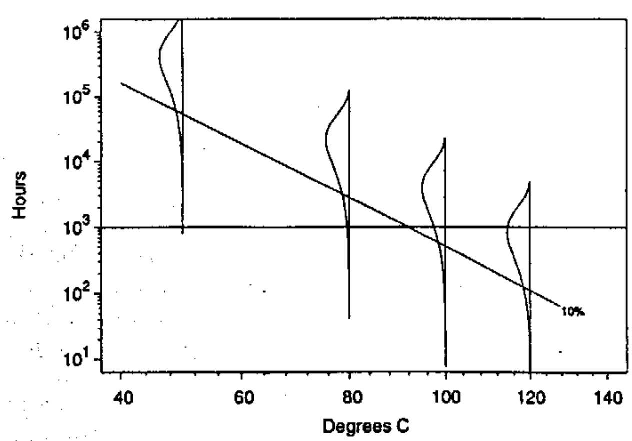

Figure 3: Possible Results for a Temperature-Accelerated Failure Mode on an IC

It is possible that such an accelerated test, while focusing on one known failure mode, might mask another! Figure 4 illustrates this, and shows that it is often the masked failure mode that is the first one to show up in the field. In such cases, the masked failure modes often dominates reported field failures.

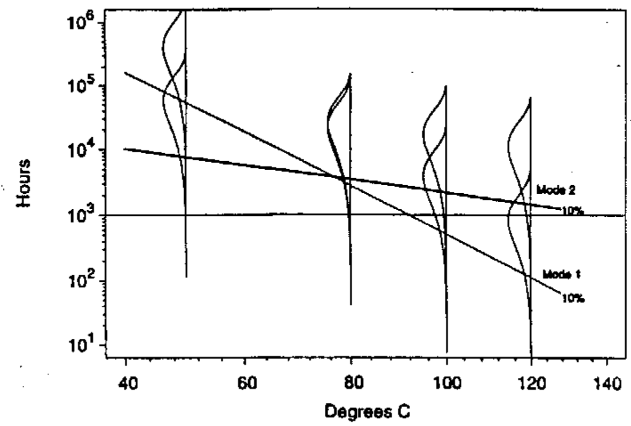

Figure 4: Unmasked Failure-Mode #2 With Lower Activation Energy

6. PITFALL 5

Faulty Comparison

It is sometimes claimed that accelerated testing is not really useful for predicting reliability, but is useful for comparing alternatives, *eg*, alternative designs, or vendors. Consider comparing similar products from two different vendors. The thought behind this claim is that laboratory accelerated tests generally cannot be expected to approximate actual use conditions adequately, but that if vendor-1 is better than vendor-2 in an accelerated test, than the same would be true in field use, as illustrated in figure 5. ALT for comparison are, however, subject to some of the same difficulties as other ALT.

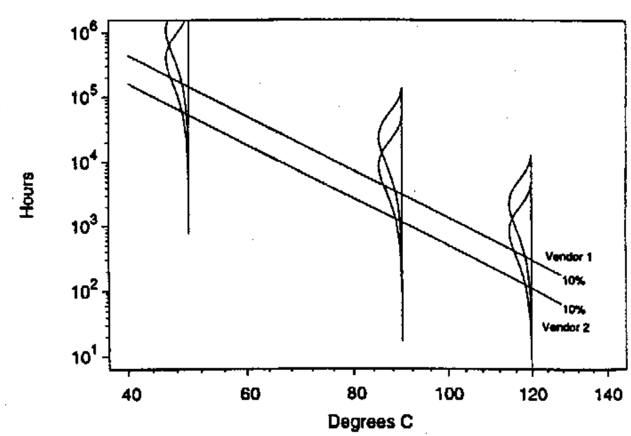

Figure 5: Well-Behaved Comparison of 2 Products

In particular, consider the results depicted in figure 6, where, vendor-1 had longer life at both of the accelerated

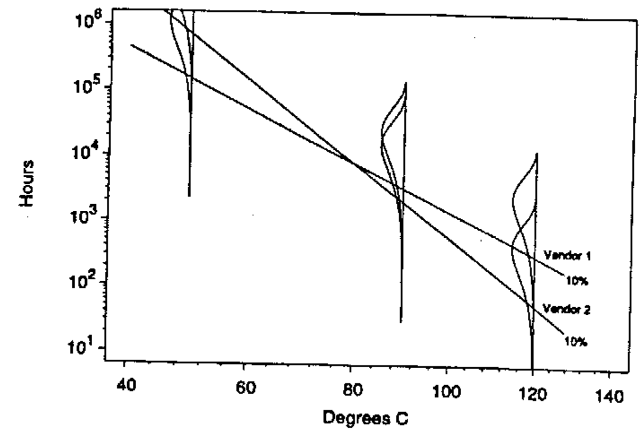

Figure 6: Comparison with Evidence of Different Failure Modes

test conditions, but the prediction at use conditions suggested that vendor-2 would have higher reliability. An important decision on the basis of limited results in this ALT would be, at best, difficult to justify. It would be most important to find out why the slopes are different and to understand the life-limiting failure modes at use conditions. If the failures at the use conditions are not the same as those at the accelerated conditions, then the ALT results would be wrong. Also, the early failures for vendor-2 might be masking the failure mode in the test results for vendor-1. One cannot, in general, use an ALT to compare products that have different kinds of failure modes.

7. PITFALL 6

Accelerating Variables Can Cause Deceleration!

In some cases it is possible that increasing what is thought to be an accelerating variable will actually cause deceleration! For example, increased temperature in an "accelerated" circuit-pack reliability audit predicted few field failures. The number of failures in the field was much higher than predicted because the increased temperature resulted in lower humidity in the "accelerated" test and the primary failure mode in the field was caused by corrosion that did not occur at high temperature and low humidity. It is for this reason that in most accelerated tests of electronic equipment, both temperature & humidity need to be controlled.

In a similar application, a higher-than-usual use-rate for a mechanical device in an accelerated test inhibited a corrosion failure mechanism. That corrosion eventually caused a serious field problem that was not predicted by the accelerated test.

In an accelerated test of a newly designed automobile air conditioner, the reliability, based on a set of constant-run accelerated life tests, was predicted to be very high over a 5-year period. However, after 2 years, a substantial fraction of the in-service air conditioners failed due to a drying-out material degradation. These failure were caused by lack of use in winter and were never seen in the continuous accelerated testing.

8. PITFALL 7

Beware of Untested Design/Production Changes

A new electro-mechanical device was to be used in a system designed for 20 years of service in a protected environment. An accelerated test of the device was conducted and this test "demonstrated" 20-year life (no more than 10% failing) under usual operating conditions (typical use rate). After the accelerated test, and as the product was going to production, a material change was made by the device vendor. The change lead to a material-degradation failure mode that caused (or would have caused) all in-service units to fail within 10 years. Eventually, all installed devices had to be replaced.

9. PITFALL 8

Beware of Drawing Conclusions on the Basis of Specially Built Prototype Test Units

Seriously incorrect conclusions can result from an accelerated life test if the test-units differ importantly from actual production-units. For example, factory manufacturing conditions are different than those in a laboratory. Cleanliness & care in building prototype *vs* production units can differ substantially. Material & parts in prototype units might differ from those that will be used in production. Highly trained technicians might build prototype units in a manner differing from, and using different tools than, the factory workers.

As much as possible, test units for an accelerated test should be taken from actual or simulated production conditions using the same raw materials, parts, *etc*, that will be used in actual production. Manufacturing processes should match, as close as possible, those that will be used in actual manufacturing to reflect variabilities that are present in actual production.

In one situation, an accelerated test was conducted on 12 prototype units. The units contained epoxy that had to be cured in an oven for a specified amount of time. The product passed its accelerated test with a safe margin. In actual manufacturing operations, however, the curing process was not well controlled. Uncured epoxy can be highly reactive. For this product, a substantial proportion of installed units eventually failed prematurely due to corrosion caused by improperly controlled curing.

10. PITFALL 9

It Is Difficult to Use Accelerated Life Tests to Predict Field Reliability

Laboratory accelerated tests are generally run at carefully controlled conditions. In some cases, however, there is unanticipated variability in experimental conditions, *eg* [3]. In many applications, field environment is anything but carefully controlled. For example outdoor weathering highly depends on the variable environmental conditions, *eg*, paints and coating subjected to UV radiation, varying temperature, humidity, as well as harmful chemical compounds in acid rain. In general, laboratory accelerated testing has not been useful for predicting the life of paints & coatings in outdoor environments. The industry still depends heavily on the use of expensive longer-term outdoor testing. Ref [4] discusses these issues. One possible reason for this difficulty is that, in general, substituting average environmental conditions into one's ALT model will NOT provide an adequate prediction of field reliability. The variability in environmental conditions itself can have a seriously detrimental effect on product reliability.

11. CONCLUDING REMARKS

Individuals involved in testing materials & components to get information on durability and life-distribution characteristics are generally seeking sensible, appropriate methods for accelerated testing. Pressure to further shorten product development cycle times is increasing the need for accelerated testing. Many instructors of elementary statistics courses make a strong, correct, statement that extrapolation is dangerous. In accelerated testing, however, all or most desired conclusions require extrapolation out-side of the range of the available data. Users and potential users of accelerated test methods must, therefore, use extreme caution. As stressed in [1], the fact that accelerated testing is the only game-in-town is not sufficient to warrant its use (and potential misuse). Appropriate use of accelerated test methods requires careful consideration (theoretical & experimental) of the underlying failure mechanisms and the effect that potential accelerating variables will have on these mechanisms. Also see [2].

ACKNOWLEDGMENT

While writing this paper, we benefitted from helpful discussions on the subject with Marci Conors, Mike LuValle, Gerry Hahn, and Wayne Nelson. Parts of this paper were adapted from Meeker & Escobar [5] with permission from John Wiley & Sons. Computing for the research reported in this paper was done, in part, using equipment purchased with funds provided by an NSF SCREMS grant award DMS 9707740 to the Department of Statistics at Iowa State University.

REFERENCES

[1] R.A. Evans, "Accelerated testing: The only game in town", *IEEE Trans. Reliability*, vol R-26, 1977 Oct, p 241.

[2] R.A. Evans, "Accelerated testing", *IEEE Trans. Reliability*, vol 40, 1991 Dec, p 497.

[3] R.M. Fischer, W.D. Ketola, W.P. Murray, "Thermal variability in accelerated weathering devices", *Progress in Organic Coatings*, vol 19, 1991, pp 165 – 179.

[4] J.W. Martin, S.C. Saunders, F.L. Floyd, J.P. Wineburg, "Methodologies for predicting the service lives of coating systems", 1996; Federation of Societies for Coatings Technology, Blue Bell.

[5] W.Q. Meeker, L.A. Escobar, *Statistical Methods for Reliability Data*, 1998; John Wiley & Sons.

[6] W. Nelson, *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses*, 1990; John Wiley & Sons.

[7] P.A. Tobias, D.C. Trindade, *Applied Reliability* (2^{nd} ed), 1995, Van Nostrand Reinhold.

AUTHORS

Dr. William Q. Meeker; Dep't of Statistics and Center for Nondestructive Evaluation; Iowa State University, Ames, Iowa 50011 USA.
Internet (e-mail): wqmeeker@iastate.edu

William Q. Meeker is a Professor of Statistics and a Distinguished Professor of Liberal Arts and Sciences at Iowa State University. He holds a BS (1972) from Clarkson College, and MS (1973) and PhD (1975) from Union College. He has consulted extensively on problems in reliability data analysis, reliability test planning, accelerated testing, and statistical computing for companies such as AT&T Bell Laboratories, General Electric Corporate Research and Development, and the Ford Motor Company. He is a Fellow of the American Statistical Association and a past Editor of *Technometrics*. He is an Associate Editor for International Statistical Review. He has won the American Society for Quality (ASQ) Youden prize and twice won the ASQ Wilcoxon Prize. He is the co-author of 3 books, 4 book-chapters, and of numerous publications in the engineering and statistical literature.

Dr. Luis A. Escobar; Dep't of Experimental Statistics; Louisiana State Univ; Baton Rouge, Louisiana 70803 USA.
Internet (e-mail): luis@stat.lsu.edu

Luis A. Escobar is a Professor in the Department of Experimental Statistics, Louisiana State University. He holds a BS in Administrative Engineering from National University, Medellin, an MS from the Inter-American Statistical Training Center (CIENES), Santiago, and a PhD in Statistics from Iowa State University. He spent one year and several summers working at AT&T Bell Laboratories and, for other companies, has worked as a consultant on many different kinds of statistical problems. His research & consulting interests include statistical analysis of reliability data, accelerated testing, survival analysis, and linear & non-linear models. Escobar is an Associate Editor for *Technometrics* and for *IIE Trans. Quality and Reliability Engineering*. He is an elected member of the International Statistics Institute, and a Fellow of the American Statistical Association.

Manuscript TR98-090 received: 1998 May 30
Responsible editor: R.A. Evans
Publisher Item Identifier S 0018-9529(98)06858-4